UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
NATIONAL EXPOSURE RESEARCH LABORATORY
CINCINNATI, OH 45268

August 13, 1996

OFFICE OF
RESEARCH AND DEVELOPMENT

Kenneth E. Osborn
East Bay MUD
P. O. Box 24055
Oakland, CA 94623

Dear Mr. Osborn:

I vote negative on the draft of Part 1030E, proposed on 7/15/96 for inclusion into the 20th Edition of STANDARD METHODS, for the following reasons:

The draft begins by defining two terms, the "critical level" (L_c) and the "detection limit" (L_D), as they were defined by Currie, and proposes statistical procedures for producing estimates appropriate to each term. My problems with these terms, as defined, are as follows:

In the Introduction, the draft says, "...when the true concentration is at the L_D the probability of a measured concentration below L_c is 1%. The detection limit L_D may be relied upon to lead to a detection in 99% of the cases..." It should be obvious that these statements assume that analytical results above L_c are reported as produced. However, by longstanding tradition, many laboratories do not report analytical results below some reporting limit, which is usually whatever they have been given as a "detection limit". If L_D is used as the reporting limit, then 50% of the analytical results from a sample with a true L_D concentration will be reported as "less than L_D", and therefore will NOT "lead to a detection." In other words, if L_D is used as a reporting limit, as it is likely to be if it is called the "detection limit", the assumptions underlying L_D are violated.

Before Currie, L_c was commonly called the "detection limit" by many authors and is still called the detection limit by many (e.g. ACS). I think it is appropriate to call L_c the detection limit because it can be used as a reporting limit without violating any of the assumptions underlying L_c or L_D. As noted on page 3 of this draft, the USEPA has defined the MDL, which is conceptually very similar to L_c, and which is also frequently used as a reporting limit.

L_D, as defined, has been frequently called the "quantitation limit." With L_c as the reporting limit and a sample at a true concentration of L_D, there is a 99% probability that the actual quantitative analytical results will be preserved and, therefore, available for interpretion regarding detection or for any quantitative interest that a data user might have.

The definition of L_D assumes that L_C is known with certainty and that a common estimate of L_C is universally used as the reporting limit. Since L_C estimates often have considerable uncertainty associated with them, usually because they were made from a limited number of analytical results, and since the L_C estimate being used may differ among laboratories, some authors have proposed more conservative "quantitation limits", i.e., higher than L_D. Examples of this are the limit of quantitation (LOQ) defined by ACS and the PQL defined by the USEPA. "Quantitation limits" relate to limits on the interpretation of data by users, e.g., regulators, and should NEVER be used as reporting limits by data producers, i.e., laboratories.

As the draft says on page 3, the true standard deviation at the MDL (σ_{MDL}) may not be equal to the true standard deviation at C (σ_C). However, since C = 3 to 5 times MDL, you should be able to assume $\sigma_C \geq \sigma_{MDL}$. Thus, if there is any bias, the MDL procedure should produce a conservative, i.e., high biased, estimate of the true MDL.

Regarding "calibration designs" (pages 3 to 6), if they are conducted "...in the range of the hypothesized detection limit..." as specified at the bottom of pages 3 and 4, I think they are indispensable as a preliminary study whenever there is little or no information available regarding the true detection limit. However, even though the second sentence of the bottom paragraph on page 4 says, "Beginning with a calibration design ... with concentrations throughout the range of L_C to L_D...," the illustrations of the calibration approach, Figures 1 to 5, all show a MUCH broader concentration range. This is a common problem with the calibration design studies I have seen conducted, and I agree with the author's statements (see page 5) regarding the undesirable influence that concentrations much above L_D can have on this approach to estimating L_C and/or L_D. I think it is better to do the calibration study over a concentration range that tops out near L_D than to try to compensate for too broad a concentration range by using any of the weighed linear regression approaches mentioned.

I see no justification for using tolerance-interval calculations as a basis for making the point estimates, L_C and L_D. Prediction-interval calculations are designed for point estimation and provide more than adequate statistical protection against these estimates being too low. Also, regarding the detection decision, it is always made relative to each individual analytical result, not to a group of results.

Regarding the discussion on pages 7 and 8:

- I agree that the analysts in studies designed to produce detection limit estimates must not know the contents or the concentration of study samples, and ideally should not know the samples are different from routine samples. One limitation; I have

no problem with the replicate samples in an MDL study being dispersed throughout a run, day, etc., so long as they are all analyzed within the same calibration of the analytical system.

- I also agree that analysts should not modify original analytical results in these studies, although, as with routine data, analysis of related quality control data would be a valid reason for not reporting specific analytical results produced when the system was judged to be "out of control."

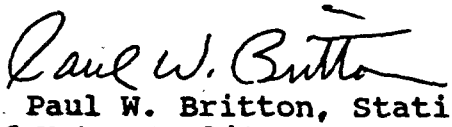
- And I agree that analyte-present experiments (like MDL studies) are better than analyte-absent experiments.

- However, I STRONGLY <u>disagree</u> with the idea of including among-instrument, among-analyst or among-laboratory components of variability into detection limit estimates. The variability relevant to each detection question is the variability of the analytical system that produced the response in question and how that relates to the average signal produced from analyses of zero-concentration samples IN THAT SYSTEM. Often the objective is to estimate the generic or average detection limit, so it is highly desirable to pool estimates of <u>single-operator</u> variability from several sets of replicate analyses, even from different laboratories, obviously the more data the better, but I believe that inclusion of other sources of variability is irrelevant and inappropriate.

What is needed in STANDARD METHODS? Is it a discussion of the various approaches to the detection problem what is needed, or is the need for a single, simple and straightforward recommendation? If the former is the objective, it still seems necessary to conclude with a clear recommendation regarding usage of the term "detection limit" and procedures for developing statistical estimates appropriate to that usage. The current draft does NOT conclude with a clear recommendation.

Once a good preliminary estimate of the detection limit is available, I think the procedures used by the USEPA to estimate the MDL, for possible use as a reporting limit, and from it the ML or PQL as "quantitation limits", are practical and reasonable approaches to recommend in STANDARD METHODS.

Sincerely yours,



Paul W. Britton, Statistician
National Water Quality Assurance Programs Br.
Ecological Exposure Research Div.

cc: William Telliard, Office of Water (4303), USEPA, Wash., DC
Henry Kahn, Office of Water (4303), USEPA, Wash., DC

STANDARD METHODS FOR THE EXAMINATION OF WATER AND WASTEWATER

JOINT TASK GROUP BALLOT

Ballot # 1

TO: Joint Task Group for Section 1030 Data Quality

ISSUE DATE: July 15, 1996 **CLOSING DATE:** August 15, 1996

RETURN BALLOT TO:

Kenneth E. Osborn
East Bay MUD
P.O. Box 24055
Oakland, CA 94623
510-287-1434
FAX: 510-465-5462

TOPIC: See enclosed copy of 1030 Data Quality

QUESTION:		**YES**	**NO**	**ABSTAIN**
	Do you approve of the enclosed document as presented?	( )	(X)	( )

Negative votes must be accompanied by a statement of the specific technical objections to the method or they will not be considered. Negative votes should be cast only when specific, substantive technical objections are stated which, if not corrected, will compromise the validity of the method. Editorial comments, correcting English or arithmetic in the method, are encouraged also, but they do not constitute the basis for a negative vote.

Comments: _____

Signed Paul W. Britton

Printed Name PAUL W. BRITTON

Date 8/13/96

c:\memos\ballot.jtg

EBMUD
LABORATORY
SERVICES

July 15, 1996

To: See Distribution List

From: Kenneth E. Osborn
(510) 287-1434 (phone)
(510) 465-5462 (fax)
KOSBORN@EBMUD.COM

Re: Draft for Standard Methods Part 1030E, Detection Limits

Please review the enclosed guidance on detection limits (part1030E) by Dr. Robert Gibbons. This is an excellent start and provides us an opportunity to develop the concept(s) of the detection limit more fully and rigorously than in the current edition of Standard Methods.

In your review, consider the audience, nature and inherent complexity of this topic, and the real need to provide a rigorous definition of the detection limit.

If you approve of the draft as written, indicate a "yes' on the accompanying ballot. If you would like to see changes or a major revision, please provide a revision of the draft.

If you have access to e-mail, please send me your revisions electronically. If you have any questions, call me at (510) 287-1434.

Sincerely,

KENNETH E. OSBORN
Quality Assurance Officer

KEO:dh

Enclosures

w:\wppublic\keo\sm1030e.wp

1030 E. DETECTION LIMIT

1. Introduction

The detection limit is a statistical estimate that is used to make the binary decision of whether or not the true concentration in a given sample is greater than zero. A frequent confusion regarding the detection limit is that measured concentrations exceeding the detection limit are quantifiable. This is clearly not the case. Measured concentrations above the detection limit only allow us to conclude that the analyte is present in the sample at a concentration greater than zero. Quantification limits have been developed for the purpose of quantitative determination[1,2].

Currie[1] defined the detection limit L_D as the true concentration "at which a given analytical procedure may be relied upon to lead to a detection." Note that the emphasis here is on "true concentration" and not measured concentration. Currie also defined the "critical level" L_C as the measured concentration "at which one may decide whether or not the result of an analysis indicates detection". It is critically important to understand the difference between L_C and L_D. When the true concentration is equal to L_C the probability of detecting it is only 50%. In contrast, when the true concentration is at the L_D the probability of a measured concentration below L_C is 1%. The detection limit L_D may be relied upon to lead to a detection in 99% of the cases, whereas the critical level does not. For this reason, the following discussion is focused primarily on estimation of the detection limit.

2. Determining the Detection Limit

There are many different names for the detection limit which only leads to further confusion. Often the problem arises due to confusion between the detection limit and the critical level, the choice of statistical multiplier, use of blanks, a single concentration or a multiple concentration calibration design. Unfortunately different investigators have given different names to the various statistical approaches to estimating the same thing, *i.e.*, the detection limit. While the distinction between the critical level and the detection limit are qualitative and are therefore deserving of different names, the other distinctions are not. In the following sections, several statistical approaches to estimating the detection limit L_D are described. The choice among these different estimators should be based on the different assumptions made by these methods, the suit-

that the risk of false positive and false negative rates is equivalent (*i.e.*, $z_{1-\alpha} = z_{1-\beta} = z$) then the detection limit is simply:

$$L_D = L_C + z_\beta \sigma_D = z(\sigma_0 + \sigma_D) = 2L_C ,$$

or twice the critical level. For $\alpha = \beta = .01$, the detection limit is therefore $4.66\sigma_0$. If the true concentration is L_D then the probability of a measured valued below L_C is 1%.

In reviewing Currie's method it is critically important to note that the only case considered is the one in which the population values σ_0 and σ_D are known. In practice, however population values are never known and the methods described in following sections are required to incorporate uncertainty in sample-based estimates of these statistics in computing detection limit estimates.

In an attempt to provide small sample properties to Currie's L_C, Glaser[6] replaced the normal distribution with Student's t-distribution and called it the "method detection limit" (MDL) *i.e.*,

$$\text{MDL} = t_{[0.01,\, n-1]} s_C ,$$

where s_C is defined as the standard deviation of n analytical replicates spiked at a single concentration C. This is an estimate of the critical level and not a detection limit therefore it is not considered further despite it's widespread use by USEPA (see Gibbons[7] for a review of statistical problems with this approach and corresponding responses by its proponents at USEPA).

The critical assumption underlying single concentration designs is that variability is homogeneous in the range of possible spiking concentrations and the true limit of detection. This assumption is implicit in use of observed spiking concentration variance as an estimator of the true variance at the limit of detection. This assumption is rarely realized in practice.

2.2 Calibration Designs

An alternative method for estimating detection limits is a calibration design. In this case, a series of samples are spiked at known concentrations in the range of the hypothesized detection limit, and variability is determined by examining the deviations of the actual response signals from the fitted regression line of response signal on known concentration. In these designs, it is generally assumed that the deviations from the fitted regression line are normally distributed.

where $s(\hat{y}_x)$ is the estimated uncertainty of a predicted instrument response (or measured concentration) $\hat{y} = b_0 + b_1 x$ at true concentration x (see Appendix). The detection limit is defined as the point at which we can have 99% confidence that the response signal is greater than L_C; therefore, Hubaux and Vos suggest that the response signal be obtained graphically by locating the abscissa corresponding to L_C on the lower prediction limit (see Figure 1). A more direct solution for L_D is provided in the technical Appendix.

The method of Hubaux and Vos assumes that variability is constant throughout the range of concentrations used in the calibration design. If this assumption is violated, then the detection limit will be overestimated because variability at high concentrations are given equal weight as those at lower concentrations. Clayton[11] suggested a variance stabilizing square root transformation which helps to some extent but does not eliminate the problem. Oppenheimer[12] proposed an estimator of L_D based on WLS regression which provides a general solution to this problem but requires an iterative solution (see Appendix). Gibbons[13] provide a noniterative computing approximate and further generalized this result to the case of multiple future detection decisions by substituting tolerance limits for prediction limits. See Gibbons[7,14] for a review of this literature.

Illustration

Figures 1-5 display actual and measured concentrations of benzene in 22 five-point concentration calibrations. Figure 1 represents a prediction interval with constant variance and corresponding Hubaux-Vos estimates of L_C and L_D. Figure 2 presents the same data assuming nonconstant variance and corresponding WLS prediction limits. Figure 3 represents the same data as Figure 2 but shows error bands based on a statistical tolerance interval with 95% confidence and 99% coverage. Figure 1 reveals that if we assume constant variance, in order to incorporate larger variability at higher concentrations, variability is overestimated at the lower concentrations. In contrast, the WLS approach described in Figure 2 provides excellent fit to the observed data for both high and low true concentrations. Figure 3 reveals that substitution of tolerance intervals for prediction intervals provides slightly wider intervals but yields detection limits that will include 99% of all future detection decisions whereas prediction limits only provide this level of confidence for a single future detection decision. Figure 4 displays the WLS approximation described by Gibbons and co-workers based on the assumption of proportional variance and concentration. Figure 4 reveals quite similar results to Figure 2 where variance is explicitly modeled as an exponential regression function of concen-

some detail by others[15,16]. There are several guiding concepts critical for producing unbiased detection limit estimates of practical relevance.

First, in analyte present studies, the analysts must be blind to both the number of compounds in the sample and to their spiking concentrations. To achieve this goal, the number of compounds must vary (perhaps randomly) from sample to sample. Furthermore, the concentration of each constituent should vary both within and across samples. Without insuring that the analyst is blind to both presence and concentration of the analyte under study, the resulting detection limit simply cannot be applied to routine practice in which such uncertainty must always exist. In practice, it is often impossible to execute such studies since numerous samples would have to be prepared at widely varying concentrations. In the absence of this level of experimental control, standard calibration data in which the analysts are unaware that they are being tested may have to suffice. The critical issue is that the analysts must *not* go back and retest samples that appear to be anomalous relative to the known spiking concentration.

Second, two or more instruments and analysts must be used and the assignment of samples to analysts and instruments must also be random. Since in large production laboratories, any one of a number of analysts and/or instruments may be called upon to analyze a test sample, this same component of variability must be included in determining the detection limit.

Third, if multiple laboratories are used or a regulatory agency analyzes split-samples or additional samples from the facility, then the entire detection limit study must be replicated in multiple laboratories. Data from a single laboratory should only be used when it is technically unfeasible to provide common calibration standards, or to split common standard samples, or a dedicated laboratory is used for all relevant analyses.

Fourth, the number of samples selected should be based on statistical power criteria, such that a reasonable balance of false positive and false negative rates is achieved. For example, if we estimate σ by computing s on seven samples, our uncertainty in σ will be extremely large and our resulting detection limit estimate L_D will also be quite large. By increasing the number of samples to say, 25, we achieve a much more reasonable estimate of σ, and resulting L_D are greatly reduced. The cost of running a few additional samples is far cheaper than dealing with the drawbacks of using detection limits incapable of detecting anything but the largest signals.

An additional note regarding analyte-absent experiments (*i.e.*, blank samples), rather than running a series of blank samples at once, they should be randomly entered into the analysts' work load throughout the course of the day. The purpose of this approach is to ensure that the analysts are blind to sample composition. The broader question is whether analyte-absent experi-

[11] Clayton. C.A., Hines, J.W. and Elkins, P.D. (1987). Detection limits with specified assurance probabilities. *Anal. Chem.*, **59**. 2506-2514.

[12] Oppenheimer, L. Capizi, T.P., Weppelman. R.M. and Mehta, H. (1983). Determining the lowest limit of reliable assay measurement. *Analytical Chemistry*, **55**, 638-643.

[13] Gibbons, R.D., Jarke, F.H. and Stoub, K.P. (1991). Detection limits for linear calibration curves with increasing variance and multiple future detection decisions. *Waste Testing and Quality Assurance*, **3, ASTM STP 1075**, D. Friedman, ed. American Society for Testing and Materials,Philadelphia, PA, 377-390.

[14] Gibbons. R.D. (1994). *Statistical Methods for Groundwater Monitoring*. Wiley, New York.

[15] Liteanu. C. and Rica, I. (1980). *Statistical Theory and Methodology of Trace Analysis*. Halsted Press, Chicago.

[16] Davis C.B. (1994). Environmental regulatory statistics. In *Handbook of Statistics: Environmental Statistics*, **12**, Chapter 26, G.P. Patil and C.R. Rao, Eds, Elsevier.

[17] Zorn, M. (1996). Detection limits with nonconstant variance. Unpublished Doctoral Dissertation. University of Wisconsin, Madison.

[18] Caulcutt, R. & Boddy, R. (1983). *Statistics for Analytical Chemists*, Chapman and Hall, New York.

[19] Gibbons. R.D. Coleman, D.E. and Maddalone, R.F. (1996). An alternative minimum level. submitted for publication.

[20] Rocke, D.M. and Lorenzato, S. (1995). A two-component model for measurement error in analytical chemistry. *Technometrics*, **37**, 176-184.

TECHNICAL APPENDIX

Details of the WLS prediction interval approach described by Oppenheimer[12] are provided in the following. To obtain the Hubaux-Vos result set all weights $k_i = 1$. To obtain a detection limit based on tolerance intervals see equation 5.3.2 in Gibbons[14] and Zorn[17].

Compute the weighted least squares regression of measured concentration or instrument response (y) on true concentration (x) for the linear model

$$\hat{y} = b_{0w} + b_{1w} x_i \,, \tag{1}$$

where

$$b_{1w} = \frac{\sum_{i=1}^{n} [(x_i - \bar{x}_w) y_i / k_i]}{\sum_{i=1}^{n} [(x_i - \bar{x}_w)^2 / k_i]} \,, \tag{2}$$

$$\bar{y}_w = \frac{\sum_{i=1}^{n} [y_i / k_i]}{\sum_{i=1}^{n} [1/k_i]} \,, \tag{3}$$

$$\bar{x}_w = \frac{\sum_{i=1}^{n} [x_i / k_i]}{\sum_{i=1}^{n} [1/k_i]} \,, \tag{4}$$

and the weight $k_i = s_{x_i}^2$ is the variance for sample i. The weighted residual variance is

$$s_w^2 = \sum_{i=1}^{n} \left[(y_i - \hat{y}_{wi})^2 / k_i \right] / (n-2) \,. \tag{5}$$

and the estimated variance for a predicted value $\hat{y}_{wj}$ is

$$V(\hat{y}_{wj}) = s_w^2 \left[k_j + \frac{1}{\sum_{i=1}^{n} (1/k_i)} + \frac{(x_j - \bar{x}_w)^2}{\sum_{i=1}^{n} (x_i - \bar{x}_w)^2 / k_i} \right] \,, \tag{6}$$

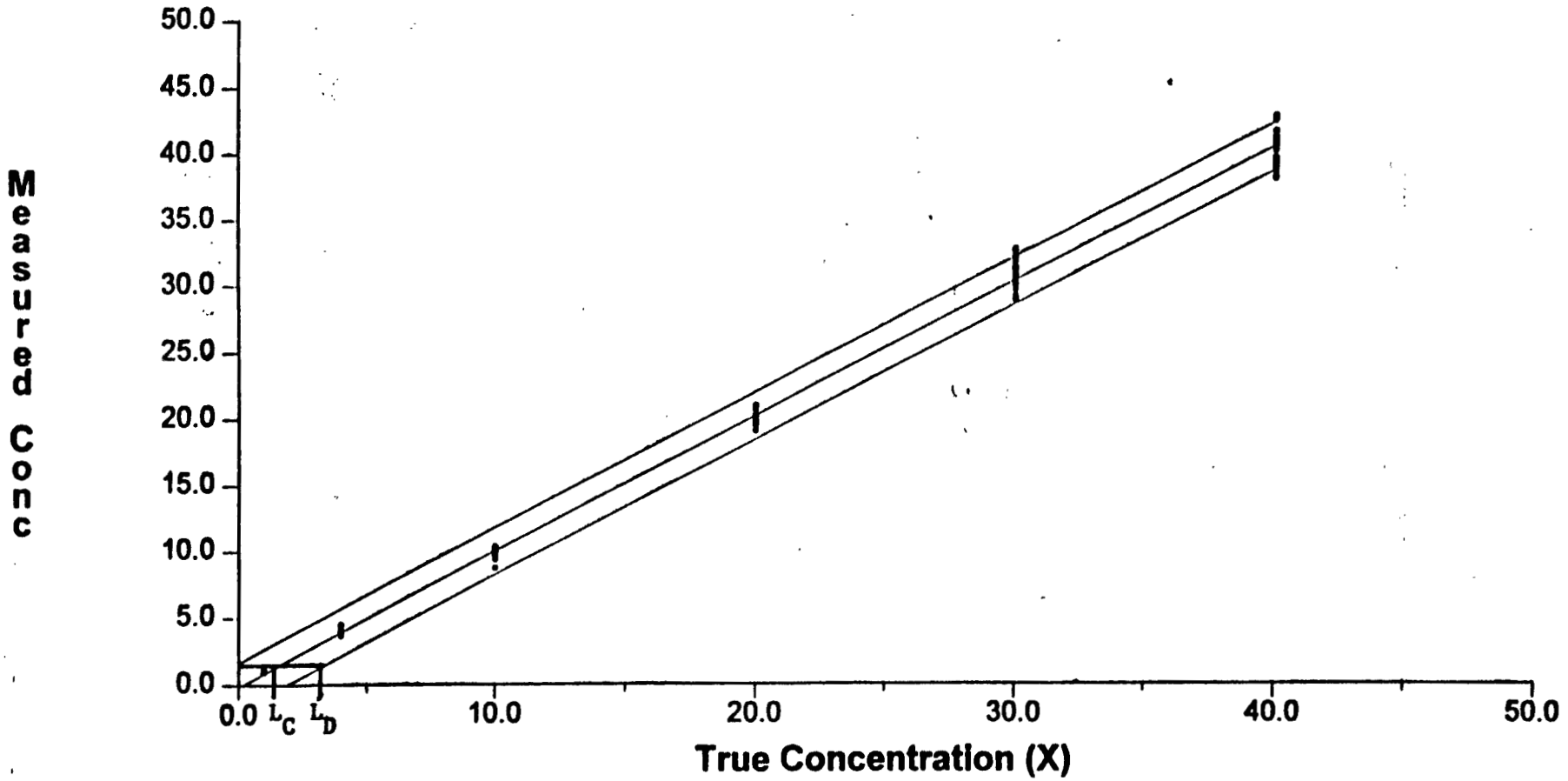
where k_j is the estimated variance at concentration x_j. An upper $(1-\alpha)100\%$ confidence interval for $\hat{y}_{wj}$ (*i.e.*, a prediction interval for a new measured concentration or instrument response at true concentration x_j) is

$$\hat{y}_{wj} + t\sqrt{V(\hat{y}_{wj})} \,, \tag{7}$$

where t is the upper $(1-\alpha)100$ percentage point of Student's t-distribution on $n-2$ degrees of freedom[12,18]. The WLS estimate of L_C is therefore

$$L_C = \frac{t s_w}{b_{1w}} \sqrt{s_{L_C}^2 + \frac{1}{\sum_{i=1}^{n} (1/k_i)} + \frac{(L_C - \bar{x}_w)^2}{\sum_{i=1}^{n} (x_i - \bar{x}_w)^2 / k_i}} \,, \tag{8}$$

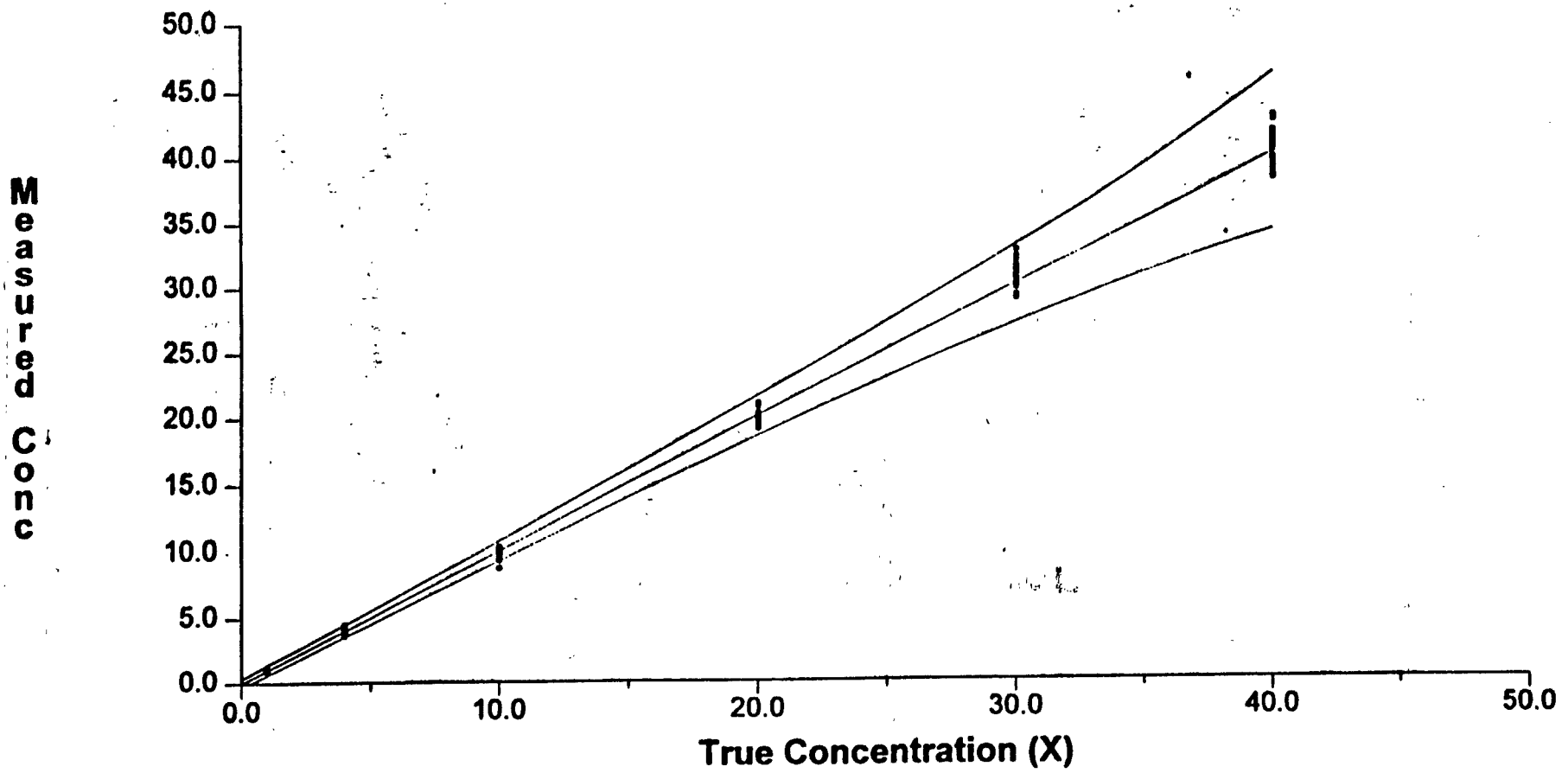
Figure 1



Critical Level (L_C) = 1.758

Detection Limit (L_D) = 3.514

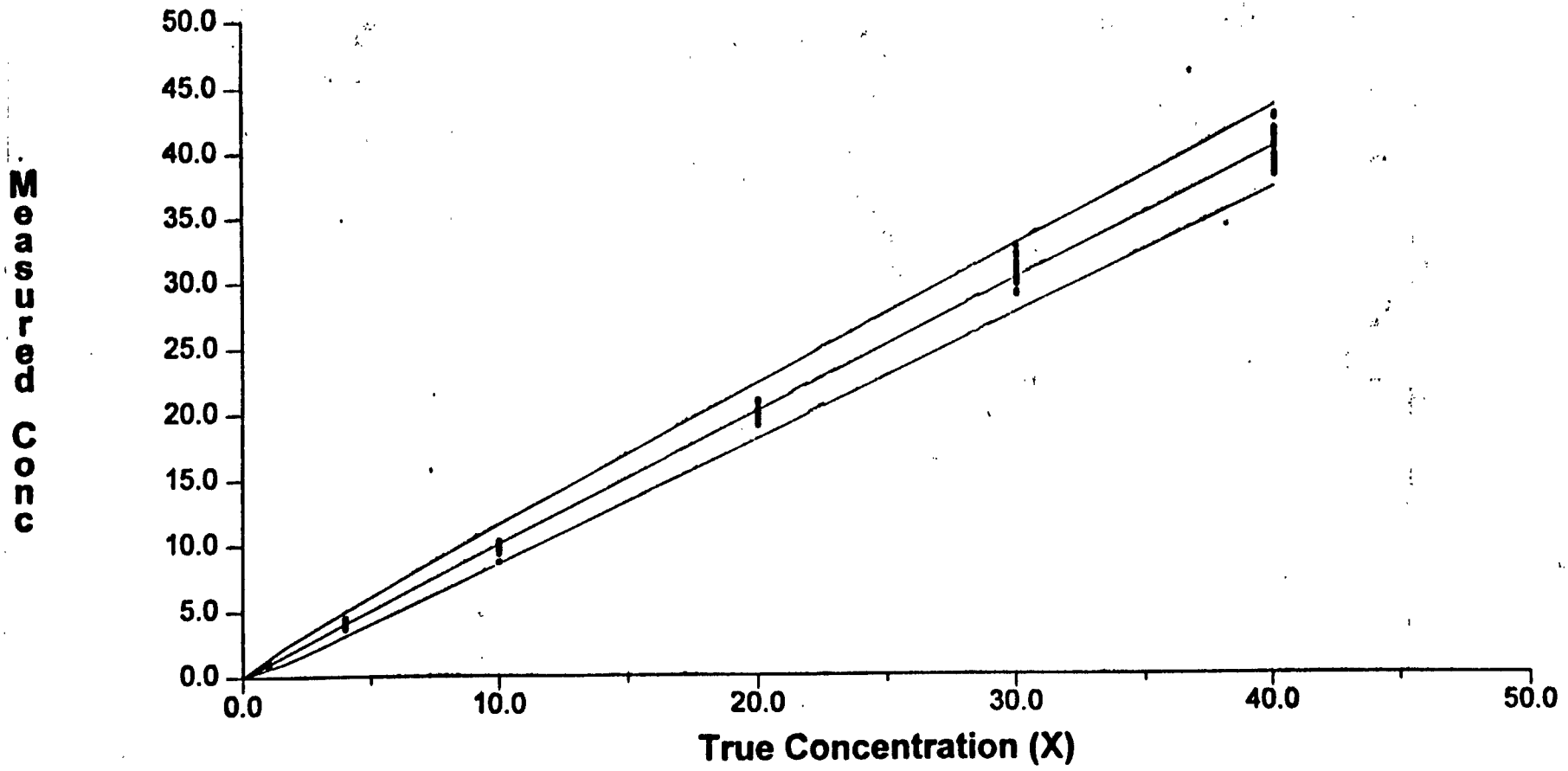
Figure 3



Critical Level (L_C) = 0.363

Detection Limit (L_D) = 0.733

Figure 5



Critical Level (L_C) = 0.463

Detection Limit (L_D) = 0.777